

University of Dundee

Clinical performance of Siemens digital breast tomosynthesis versus standard supplementary mammography for the assessment of screen-detected soft-tissue abnormalities – a multi-reader study

Whelehan, P.; Heywang-Köbrunner, S. H.; Vinnicombe, S. J.; Hacker, A.; Jaensch, A.; Hapca, A.

Published in:
Clinical Radiology

DOI:
[10.1016/j.crad.2016.08.011](https://doi.org/10.1016/j.crad.2016.08.011)

Publication date:
2017

Licence:
CC BY-NC-ND

Document Version
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Whelehan, P., Heywang-Köbrunner, S. H., Vinnicombe, S. J., Hacker, A., Jaensch, A., Hapca, A., Gray, R., Jenkin, M., Lowry, K., Oeppen, R., Reilly, M., Stahnke, M., & Evans, A. (2017). Clinical performance of Siemens digital breast tomosynthesis versus standard supplementary mammography for the assessment of screen-detected soft-tissue abnormalities – a multi-reader study. *Clinical Radiology*, 72(1), 95.e9-95.e15. <https://doi.org/10.1016/j.crad.2016.08.011>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Clinical performance of Siemens digital breast tomosynthesis versus standard supplementary mammography for the assessment of screen-detected soft-tissue abnormalities – a multi-reader study

Patsy Whelehan¹, Sylvia H. Heywang-Köbrunner², Sarah Vinnicombe¹, Astrid Hacker², Alexander Jaensch², Adrian Hapca¹, Rosie Gray³, Maggie Jenkin⁴, Keith Lowry⁵, Rachel Oeppen⁶, Michael Reilly⁷, Michaela Stahnke⁶, Andy Evans

1. School of Medicine, University of Dundee, Mailbox 4, Ninewells Hospital & Medical School, Dundee, DD1 9SY, UK

Corresponding author: p.j.whelehan@dundee.ac.uk 01382 383292

a.z.evans@dundee.ac.uk

S.vinnicombe@dundee.ac.uk

2. Referenzzentrum Mammographie München, Sonnenstraße 29, 80331 München, Germany

3. Royal Cornwall Hospital, Truro, UK

4. Derriford Hospital, Plymouth, UK

5. Belfast City Hospital, UK

6. Princess Anne Hospital, Southampton, UK

7. Altnagelvin Hospital, Londonderry, UK

Acknowledgments

Thanks are due to Dr Kulsam Ali and Dr Violet Warwick for data entry and data management.

This study was funded by Siemens Healthcare Ltd. The funder had no input into the study design, analysis or reporting.

1 guarantor of integrity of the entire study

Andy Evans

2 study concepts and design

Andy Evans, Sylvia Heywang-Koebrunner, Patsy Whelehan, Sarah Vinnicombe, Adrian Hapca

3 literature research

Andy Evans, Patsy Whelehan

4 clinical studies

Patsy Whelehan, Sylvia Heywang-Koebrunner, Sarah Vinnicombe, Astrid Hacker, Alexander Jaensch, Rosie Gray, Maggie Jenkin, Keith Lowry, Rachel Oeppen, Michael Reilly, Michaela Stahnke, Andy Evans

5 experimental studies / data analysis

Adrian Hapca, Patsy Whelehan

6 statistical analysis

Adrian Hapca, Patsy Whelehan

7 manuscript preparation

Andy Evans, Sylvia Heywang-Koebrunner, Patsy Whelehan

8 manuscript editing

Patsy Whelehan, Sylvia Heywang-Koebrunner, Sarah Vinnicombe, Astrid Hacker, Alexander Jaensch, Adrian Hapca, Rosie Gray, Maggie Jenkin, Keith Lowry, Rachel Oeppen, Michael Reilly, Michaela Stahnke, Andy Evans

- Single-view Siemens DBT has equivalent accuracy to supplementary mammographic assessment.
- The auROC curve was 0.870 for DBT and 0.857 for standard supplementary mammography ($p=0.4890$).
- DBT allows better detection of extra cancer foci compared to standard supplementary mammography.

Abstract

Objectives

We aimed to compare the diagnostic accuracy of standard screening images plus single-view DBT - using Siemens DBT equipment - with standard screening images plus supplementary mammographic views in non-calcific, screen detected mammographic abnormalities.

Methods

Participants were unselected women aged 50-69 recalled within a population-based European breast screening programme for assessment of soft-tissue mammographic abnormalities. Supplementary mammographic views (SMVs) and DBT were performed in all cases. A range of equipment was used for screening and supplementary mammography but all DBT examinations were performed using the Siemens Mammomat Inspiration. A retrospective multi-reader study including 238 cases for whom either histology or at least two years' follow-up was available was performed with eight suitably accredited UK breast screening personnel reading all cases under both conditions, with temporal separation. Readers were blinded to case outcomes and findings from other examinations. Diagnostic accuracy using Receiver Operating Characteristic analysis was compared between screening plus SMV images and screening plus DBT images. The study was powered to detect a 3% inferiority margin in diagnostic accuracy between methods.

Results

The final sample with complete data available for analysis included 195 benign cases (1,560 reads) and 35 malignant cases (280 reads). The DBT method yielded a slightly higher AUC value than the SMV method (0.870 vs 0.857) but the difference is not statistically significant ($p=0.4890$), indicating that the methods have equivalent accuracy.

Conclusions

Siemens DBT demonstrates equivalent diagnostic accuracy according to ROC curve analysis when used in place of SMVs in screen-detected soft tissue mammographic abnormalities.

Introduction

Supplementary mammography such as spot compression views has traditionally been performed when standard screening mammography has detected masses, distortions and asymmetric densities which do not display definitively benign appearances. While spot compression views can accurately characterise the borders of masses and confirm the presence or absence of distortion, they are sometimes wrongly interpreted as benign or normal^{1, 2}. Some of the errors arise from the lesion being displaced out of the field by the compression paddle². The limitations of spot compression views have led to the recommendation in UK assessment guidelines that ultrasound examinations should always be performed in cases of recall for asymmetric density³.

Digital breast tomosynthesis (DBT) is a recent development of full field digital mammography (FFDM). By acquiring multiple low-radiation-dose images across a range of angles and reconstructing the data to present a series of thin “slice” images, DBT ameliorates the problem of feature obscuration by overlying tissues. It has been shown that clinical accuracy can be improved as a result⁴, with DBT being particularly effective at detecting mammographic spiculation⁵. As DBT images the whole breast, displacement of the lesion out of the imaging field is not an issue and DBT also has the potential to enable the detection of additional ipsilateral lesions not apparent on the original FFDM images. These potential advantages have led to a number of studies comparing supplementary mammographic views (SMVs) and DBT in women recalled from mammographic screening for non-calcific abnormalities. These studies have tended to show that DBT is not inferior to SMVs or, in some cases, slightly superior⁶⁻¹¹. One of the larger studies, from the UK, demonstrated that for standard FFDM plus SMVs, the area under the Receiver Operating Characteristic (ROC) curve (the AUC) was 0.87, while for FFDM plus single-view DBT, the AUC was 0.93 ($p = 0.0014$)⁸. Such studies have until recently been performed using equipment from a single vendor, Hologic.

DBT equipment from other vendors is now commercially available in the UK but the technical specifications vary to such an extent that the National Health Service Breast Screening Programme (NHSBSP) in the UK requires vendor-specific data before approving the replacement of SMVs with DBT for the assessment of non-calcific screen detected abnormalities. **The specifications of the three systems which have received NHSBSP technical evaluations to date are described in a series of reports¹²⁻¹⁴. Tomosynthesis-specific variations between these systems include whether a grid is used – no in Siemens and Hologic, yes in GE; the height of the centre of rotation of the tube in relation to the detector surface; whether the projection images are obtained while the tube is in motion (Hologic and Siemens), or with a step-and-shoot process (GE); the reconstruction algorithm (filtered back projection in Siemens, filtered back projection with iterative optimisation in Hologic, iterative in**

GE); the pixel size of the reconstructed images (85 microns in Siemens, 100 microns in Hologic and GE); and the number of projection images and the angular range over which they are obtained. In the Siemens equipment, 25 projection images are obtained across 50 degrees¹⁴, in the Hologic, 15 are obtained across 15 degrees¹², and in the GE, 9 are obtained across 25 degrees¹³. In general, variations in such parameters not only interact but, in themselves, yield gains and losses in image quality and, in some cases, radiation dose. Therefore the design details always involve a trade-off. For example, a wider angular range is expected to enhance in-depth resolution, which might aid radiological interpretation, but this comes at the cost of lower in-plane resolution¹⁵.

In a recent study using GE equipment, retrospective analysis with blinding to the opposite condition demonstrated that the performance of two-view screening FFDM plus two-view DBT performed at assessment was non-inferior to the performance of two-view FFDM plus SMVs – AUC = 0.873 (95% CI 0.834-0.906) and 0.900 (95% CI 0.864 -0.929) respectively, $p=0.17$ ¹⁶.

The aim of our study was to compare the diagnostic accuracy of standard screening images plus single-view assessment DBT - using Siemens DBT equipment - with standard screening images plus supplementary mammographic views in non-calcific, screen detected mammographic abnormalities.

Materials and Methods

This study was a collaboration between a breast screening programme in Germany and a research group in the UK. This approach was used to avoid unnecessarily repeating a prospective interventional study while at the same time providing data applicable to the UK breast screening programmes.

The study cases were acquired prospectively in a population-based screening programme in Germany, in which women aged 50-69 are invited biennially and which strictly follows European guidelines for mammography screening. During the recruitment period (May 2010 to November 2011), single-view DBT was offered to consecutive women in whom SMVs were indicated to assess a screen-detected abnormality. DBT was performed after and in addition to standard imaging assessment entailing one or more SMVs and sonographic examination. The study was approved by the local ethics committee and by the Bureau of Radiation Protection. All participants signed informed consent.

Until 1 October 2010, the initial screening mammograms had been performed using film screen mammography. After 1 October 2010, screening mammograms were exclusively performed using FFDM. The SMVs were performed using one of three available mammography units: Siemens

Mammomat Inspiration (direct radiography), Sectra MDM (direct radiography) or Hologic Lorad M IV with Kodak/Carestream DirectView 975 (computed radiography). All equipment underwent the full quality assurance protocol according to European Guidelines¹⁷ and met the required standards. DBT was performed on a Siemens Mammomat Inspiration. This system acquires 25 projections over a tomosynthesis angle of 50 degrees and reconstructs 1 mm-thick planes, or “slices”.

Outcomes of all malignant **lesions and lesions of uncertain malignant potential (pathology classification “B3”)**, and 25 of the benign lesions, were confirmed by histology. All remaining outcomes were confirmed by a minimum of two years’ follow-up. For the purpose of this multi-reader study, only cases without suspicious microcalcification as the primary reason for recall were included. Four cases were lost to follow-up and were therefore excluded. This yielded a pool of 238 cases, with 42 malignant lesions in 35 women. The mean age of women in the sample was 57.6 years (median 57, range 50-70).

The multi-reader study was designed to detect an inferiority margin of 3%. It was calculated that, with the available number of cases, eight readers would provide sufficient reads to detect a 3% difference in accuracy between the modalities, with 90% power and a one-sided alpha of 0.05.

All readers were accredited and practicing within one of the national, population-based breast screening programmes in England, Scotland or Northern Ireland. They were compliant with quality guidelines to read at least 5,000 mammograms per year and to take part regularly in assessment clinics. The readers had between one and 15 years’ experience in breast screening (**mean: 10.9 years, median 9.5 years**) and had all attended recognised hands-on short courses in breast tomosynthesis interpretation. Their real-life clinical experience with DBT ranged from zero to four years (**mean: 1.5 years, median: 0.7 years**).

Cases were presented to the readers as two sets of images: Set A included the screening mammogram and the single-view DBT; Set B included the screening mammogram and all supplementary views. Each reader read all cases once using set A and once using set B. The interval between the two readings was 9 weeks and the reading condition order was reversed for half the cases, both per-reader and overall. Ultrasound images and findings from examinations carried out during the assessment were not available to the readers.

Readers were asked to classify lesions in two ways, firstly dichotomously as “biopsy required: Yes/No” and secondly on a 0-10 scale of malignancy suspicion. Receiver operator characteristic (ROC) curves were produced with both the binary and scalar data included in the model, and the AUC values were compared using the Chi-Squared test. Sensitivity and specificity were computed

from the dichotomous classifications and compared using McNemar's test for correlated proportions.

Readers were also asked to measure the lesions they considered suspicious for cancer, using the digital measurement tools on the reporting workstations. For unifocal lesions with malignant outcomes, maximum diameter (mm) in any dimension was recorded under each reading condition and compared with the histological maximum diameter following excision. Cases treated with neoadjuvant chemotherapy were excluded. For multifocal cancers, we did not assess measurement accuracy but instead calculated the proportions of the malignant foci detected by each method.

Results

From the sample of 238 cases, eight were excluded on account of missing data from the readers; all were benign. To eliminate uncertainties in matching all the reader-identified features to each other and to the ground truth data in cases of multiple unilateral lesions, these were collapsed for the analysis such that each breast was classified as either benign or malignant. This left 230 cases (breasts) for analysis, 195 benign and 35 malignant. As each case was read by all eight readers, in total, under each condition, there were 1,560 reads of benign cases and 280 reads of malignant cases available for analysis.

The dominant radiological feature overall was mass in 83 of 230 cases (36%), asymmetric density in 80 (35%), and architectural distortion in 67 (29%). In the malignant cases, 20 of 35 were masses, 4 were asymmetric density, and 11 were architectural distortion. All cancers were unilateral. The tumour types were: invasive ductal carcinoma no special type, $n=18$ (6 each of Grade 1, 2 and 3); invasive lobular, $n=14$ (4 Grade 1, 8 Grade 2, 2 Grade 3); mixed ductal-lobular, $n=1$, (Grade 1); tubular, $n=1$ (Grade 1), DCIS only, $n=2$. Invasive tumour size was less than 15 mm in 25 cases and greater than or equal to 15 mm in 10.

Results are presented for performance of the two imaging methods across all readings per method.

ROC analysis

The ROC curves showing the combined sensitivity and specificity (accuracy) of each method are provided in Figure 1. While the DBT method has a slightly higher AUC value than the SMV method (0.870 vs 0.857), the difference is not statistically significant ($p=0.4890$), indicating that the methods have equivalent accuracy.

Sensitivity and specificity

Based on the dichotomous classification by the readers, sensitivity for cancer was 90% for the DBT method and 86% for the SMV method ($p=0.10$). Specificity was 59% for the DBT method and 64% for the SMV method ($p=0.0002$).

Detection of multiple malignant foci

Of the 35 malignant cases, four were multifocal with a total of 10 detectable foci. Of the 80 opportunities (10 foci x 8 readers) under each condition for detecting a malignant focus in these cases, 39 (49%) were detected with SMVs and 50 (63%) with DBT. Using McNemar's test for correlated proportions, the result was statistically significant: $p=0.007$. However, this test likely does not account for all aspects of the correlation between the samples and it must be remembered that the actual number of cases analysed was very small.

Measurement accuracy in unifocal malignant lesions

For unifocal cancer cases not treated with NACT the number with complete data on size were 220 under the SMV condition and 236 under the DBT condition. The disparity in numbers is largely a result of the lower sensitivity in the SMV condition; if a lesion was not identified by the reader as malignant, it was not measured. The numbers where the measurement was accurate to within 5 mm and the numbers where size was overestimated and underestimated by at least 5 mm, compared to the whole tumour diameter on histology, are shown in Table 1. This is a preliminary description only, because rigorous statistical comparison of this outcome, taking into account all the relevant complexities, is currently not resourced.

Discussion

This retrospective multi-reader study has shown non-inferiority for overall diagnostic accuracy assessed by ROC analysis when substituting single-view Siemens DBT for supplementary mammographic views in the work-up of screen-detected soft tissue mammographic abnormalities. This is broadly in line with comparable work, although published studies on this question using other manufacturers' equipment have often been too small to detect statistically significant differences^{6, 7, 9, 10}. Statistically significant improvements in diagnostic classification of soft tissue lesions with DBT have, however, been demonstrated on occasion^{8, 15}.

The AUC values in our study (0.857 and 0.870 for SMVs and DBT respectively, $p=0.4890$) were slightly smaller for both conditions than demonstrated in similar UK studies. Cornford and colleagues, using GE DBT in two views, reported AUCs of 0.873 (SMVs) and 0.900 (DBT) ($p=0.1702$)¹⁶,

while Morel and colleagues, using Hologic DBT in one view, reported 0.90 and 0.97 respectively ($p=0.005$)⁸. In studies such as this, performance can vary according to reader characteristics, lesion characteristics, and other design factors, discussed further below. Our design incorporated multiple readers for each case, whereas Cornford's¹⁶ and Morel's⁸ studies employed a single reader for each case, drawn from groups where the minimum clinical experience with tomosynthesis was 6 months. The size profiles of the tumours are not reported in these two studies, which included 113 and 74 cancers respectively, compared with 35 cancers (resulting in 280 cancer-reads per condition) in ours.

Comparing sensitivity and specificity across studies is hampered by methodological differences and by studies being underpowered to detect differences when sensitivity and specificity are considered separately. Our readers were required to use a radiological classification which was dichotomous at the outset, i.e. to state whether a lesion required a biopsy or not. Although we also collected suspicion level, this was on a different scale to the other UK studies and was not used to analyse sensitivity and specificity separately. It is possible that the biopsy/no biopsy classification was not ideal because readers, unaware of ultrasound findings, may have assigned a biopsy classification to benign-appearing (M2) masses on the assumption that they would require biopsy if solid. Nonetheless, the category definition was the same across both methods so should not have affected the differences between the two methods. The ROC analysis would have been less sensitive to any effects of this classification method on the analysis because data from the 0-10 malignancy suspicion scale were included in the model, along with the dichotomous data, and ROC analysis reflects the sensitivity/specificity trade-off.

Like Cornford and Morel, we have shown descriptive improvements in sensitivity when DBT is used in place of SMV's for assessment of soft tissue lesions but none of the studies shows a statistical difference. Sensitivities for the SMV condition and the DBT condition respectively were 86.4% and 89.6% in our study, compared with 92.0% and 93.8% in Cornford's¹⁶, and 94.6% and 100% in Morel's⁸. Lesion size was not reported in the latter two publications but in our study, lesion size was less than 15 mm in 25 of 35 malignant cases.

Specificity in our study was statistically significantly lower when single-view Siemens DBT was substituted for supplementary mammographic views for the work-up of screen-detected soft tissue abnormalities – 59.0% with DBT and 64.4% with SMVs, $p=0.0002$ (Cornford: 72.3% and 70.6% [$p=0.685$]; Morel 93.4% and 92.0% [not statistically tested]). While the descriptive specificity improvements with DBT in the latter two studies were not statistically robust, it is noteworthy that they are in the opposite direction to ours, and that specificity in our study is lower than the others regardless of condition. We could speculate that the “biopsy/no biopsy” classification may have

depressed specificity in our study, and that some low levels of experience among the readers may have affected confidence to state “no biopsy”. It is well recognised that the sensitivity-specificity trade-off varies between readers¹⁸. A Swedish trial of one-view Siemens DBT versus FFDM in the primary screening context has found lower specificity alongside higher sensitivity, particularly in the early stages of the study when reader experience was lower¹⁹. The Oslo screening trial, which tested the screening condition only and used Hologic DBT, found higher specificity for DBT plus FFDM compared with FFDM alone but only for 5 of the 8 radiologists in the study²⁰. In the context of screening assessment, it can be argued that the relative importance of sensitivity over specificity is greater than in the primary screening context because the rate of cancer in the recalled population is so much higher than the screened population. Also, a missed cancer following recall for assessment can be a terrible experience for patients and staff, undermining the patient’s confidence in the breast team and risking medico-legal action, as well as affecting prognosis. However, benign needle biopsies are harmful to women, may deter future attendance for screening²¹, and use precious resources. The size of the reduction in the risk of a false negative assessment which would be considered sufficient to justify any additional biopsies arising from lower specificity is open to debate.

Although some of our readers were inexperienced with DBT, others had considerable experience, which could be considered both a strength and a limitation of the study; a strength because the range is likely to reflect the real-world disparity of experience among the UK radiology population and a limitation because the inclusion of personnel who were not at all experienced with DBT could have caused underestimation of what might represent “typical” DBT performance. A small retrospective cancer detection study has shown a lack of difference in relative performance of 2-view Hologic DBT plus FFDM compared to FFDM according to reader experience²², while a similar study with 1-view Siemens DBT demonstrated diagnostic improvement between modalities for the least experienced reader only when benign lesions were excluded, because the reader had high false positive rates across modalities²³. The same study also showed a lack of overall performance difference between modalities for the most experienced reader. We propose to conduct further analysis of our data to investigate any differences in the relative performance of the two tests according to reader characteristics.

The main strengths of our study include the large number of reader observations, maximising the power of the study despite the modest number of cancer cases. There was effective reader blinding between conditions, and risk of recall bias was minimised by a generous temporal separation between the two reading conditions, compared with other similar studies^{8, 16}. The power calculation

for the study was based on the overall accuracy outcome so, although the ROC accuracy analysis is robust, the relatively low number of cancer cases may have prevented the detection of a statistically significant difference in sensitivity between methods. In common with other studies of DBT in the assessment context, the number of actual cancers was too low to justify subanalysis by lesion type, grade, age of participants, or breast density. With regard to breast density, screening and simulated screening studies to date have not consistently shown differential accuracy improvements across a breast density range²⁴ but a recent large study with Hologic DBT plus FFDM for screening has shown the accuracy improvement over FFDM alone to be greatest in heterogeneously dense breasts²⁵.

Although we analysed at breast level rather than at lesion level for the main outcomes, we also separately considered detection of multifocality. This was, however, based on only four cases so, although the sensitivity for multifocality was higher with DBT, the statistical difference we detected is tentative. However, our finding is broadly in line with a study by Mercier and colleagues²⁶ in which detection of multifocality by 2-view Siemens DBT was significantly better than FFDM. Given that DBT examines the whole breast whereas SMVs generally do not, better detection of additional areas of disease using DBT is only to be expected.

Descriptive data on malignant lesion measurement accuracy demonstrated a greater tendency for size underestimation in the SMV condition compared to the DBT condition, and a lower frequency overall of measurement accuracy to within 5 mm. However, DBT showed a higher frequency of overestimation. Previous authors have found statistically significant improvements in malignant lesion measurement accuracy with GE DBT compared with FFDM^{27,28}, while Mercier has found that Siemens DBT overestimated tumour size unless spicules were excluded from the measurements²⁶. We propose to perform a more formal comparison of tumour measurement accuracy alongside our planned analysis of performance according to reader characteristics.

Given the ability of DBT to reduce the problem of overlapping tissues on the image, it might be expected that producing images in two projections would not be necessary. The evidence to date is inconclusive on this point. Haq and colleagues investigated the question with Hologic DBT and found that in 31 malignant soft-tissue lesions seen on only one view at primary screening, DBT demonstrated the lesion on the same or on both views at assessment and never on the opposite view alone, indicating that a single DBT view in the screening projection more clearly demonstrating the lesion which prompted recall might be sufficient²⁹. However, the study was not large enough to be conclusive. Furthermore, generalisability to other manufacturers' equipment is restricted by the very different image acquisition and reconstruction algorithms. The Swedish trial of Siemens DBT versus FFDM for primary screening has demonstrated a considerable and statistically significant

increase in cancer detection rate with a single view³⁰. Our study achieved equivalent ROC performance when exchanging SMVs for single-view DBT in the diagnostic work-up of screen detected soft tissue abnormalities and therefore adds to the evidence that a single DBT view may be sufficient in this context. However, we cannot say how the addition of a second view would have affected the accuracy of DBT and our study did not include an assessment of radiation doses. In assessment as opposed to screening, however, the relative importance of accuracy over radiation dose is greater because of the greater frequency of cancer. Tagliafico compared radiation doses between DBT (presumed two-view) and FFDM plus SMVs using Hologic equipment and found the latter to be higher but did not report the comparison between DBT and SMVs alone¹⁰. The use of one rather than two views might be expected to limit the accuracy of lesion size estimation but our preliminary findings on measurement accuracy do not support this.

In conclusion, in common with two other manufacturers' equipment, Siemens DBT demonstrates equivalent diagnostic accuracy according to ROC curve analysis when used in place of SMVs in screen-detected soft tissue mammographic abnormalities. Consideration should therefore be given to allowing this as an alternative in screening assessment clinics, subject to satisfactory practical evaluation. Given that equivalent diagnostic accuracy was achieved with a single DBT view in our study, the need for DBT always to be performed in two views in the assessment context could also be re-evaluated, at least in the case of wide-angle tomosynthesis.

References

1. Burrell HC, Evans AJ, Robin A, Wilson M, Pinder SE. False-negative breast screening assessment. What lessons can we learn? *Clinical Radiology*. 2001; 56(5): 385-8.
2. Warren R, Allgood P, Hunnam G, Godward S, Duffy S, Programme TEABS. An audit of assessment procedures in women who develop breast cancer after a negative result. *Journal of Medical Screening*. 2004; 11(4): 180-6.
3. Liston J, Wilson AR, (eds). *Clinical Guidelines for Breast Cancer Screening Assessment* (3rd edition). Sheffield: NHS Cancer Screening Programmes; 2010.
4. Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *Breast*. 2013; 22(2): 101-8.
5. Ray KM, Turner E, Sickles EA, Joe BN. Suspicious findings at digital breast tomosynthesis occult to conventional digital mammography: Imaging features and pathology findings. *Breast Journal*. 2015; 21(5): 538-42.
6. Brandt KR, Craig DA, Hoskins TL, Henrichsen TL, Bendel EC, Brandt SR, et al. Can digital breast tomosynthesis replace conventional diagnostic mammography views for screening recalls without calcifications? A comparison study in a simulated clinical setting. *American Journal of Roentgenology*. 2013; 200(2): 291-8.
7. Hakim CM, Chough DM, Ganott MA, Sumkin JH, Zuley ML, Gur D. Digital breast tomosynthesis in the diagnostic environment: A subjective side-by-side review. *American Journal of Roentgenology*. 2010; 195(2): W172-W6.
8. Morel JC, Iqbal A, Wasan RK, Peacock C, Evans DR, Rahim R, et al. The accuracy of digital breast tomosynthesis compared with coned compression magnification mammography in the assessment of abnormalities found on mammography. *Clinical radiology*. 2014; 69(11): 1112-6.
9. Noroozian M, Hadjiiski L, Rahnama-Moghadam S, Klein KA, Jeffries DO, Pinsky RW, et al. Digital breast tomosynthesis is comparable to mammographic spot views for mass characterization. *Radiology*. 2012; 262(1): 61-8.
10. Tagliafico A, Astengo D, Cavagnetto F, Rosasco R, Rescinito G, Monetti F, et al. One-to-one comparison between digital spot compression view and digital breast tomosynthesis. *European Radiology*. 2012; 22(3): 539-44.
11. Zuley ML, Bandos AI, Ganott MA, Sumkin JH, Kelly AE, Catullo VJ, et al. Digital breast tomosynthesis versus supplemental diagnostic mammographic views for evaluation of noncalcified breast lesions. *Radiology*. 2013; 266(1): 89-95.
12. Strudley C, Looney P, Young K. Technical evaluation of Hologic Selenia Dimensions digital breast tomosynthesis system: NHSBSP Equipment Report 1307; Version 2. London; 2014.
13. Strudley C, Oduko J, Young K. Technical evaluation of GE Healthcare SenoClaire digital breast tomosynthesis system: NHS Breast Screening Programme Equipment Report 1404. London; 2016.
14. Strudley C, Warren L, Young K. Technical evaluation of Siemens Mammomat Inspiration digital breast tomosynthesis system: NHSBSP Equipment Report 1306, version 2. London; 2015.
15. Rodríguez-Ruiz A, Castillo M, Garayoa J, Chevalier M. Evaluation of the technical performance of three different commercial digital breast tomosynthesis systems in the clinical environment. *Physica Medica*. 2016; 32(6): 767-77.
16. Cornford EJ, Turnbull AE, James JJ, Tsang R, Akram T, Burrell HC, et al. Accuracy of GE digital breast tomosynthesis vs supplementary mammographic views for diagnosis of screen-detected soft-tissue breast lesions. *The British Journal of Radiology*. 2015; 89(1058).
17. Perry N, Broeders M, de Wolf C, Tornberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis (4th edition). Luxembourg: European Commission; 2006.
18. Svahn TM, Macaskill P, Houssami N. Radiologists' interpretive efficiency and variability in true- and false-positive detection when screen-reading with tomosynthesis (3D-mammography) relative to standard mammography in population screening. *Breast*. 2015; 24(6): 687-93.

19. Lång K, Nergården M, Andersson I, Rosso A, Zackrisson S. False positives in breast cancer screening with one-view breast tomosynthesis: An analysis of findings leading to recall, work-up and biopsy rates in the Malmö Breast Tomosynthesis Screening Trial. *European Radiology*. 2016; 1-9.
20. Skaane P, Bandos A, Gullien R, Eben E, Ekseth U, Haakenaasen U, et al. Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosynthesis in a Population-based Screening Program. *Radiology*. 2013; 267(1): 47-56.
21. Maxwell AJ, Beattie C, Lavelle J, Lyburn I, Sinnatamby R, Garnett S, et al. The effect of false positive breast screening examinations on subsequent attendance: Retrospective cohort study. *Journal of Medical Screening*. 2013; 20(2): 91-8.
22. Alakhras MM, Brennan PC, Rickard M, Bourne R, Mello-Thoms C. Effect of radiologists' experience on breast cancer detection and localization using digital breast tomosynthesis. *European Radiology*. 2014; 25(2): 402-9.
23. Thomassin-Naggara I, Perrot N, Dechoux S, Ribeiro C, Chopier J, De Bazelaire C. Added value of one-view breast tomosynthesis combined with digital mammography according to reader experience. *European Journal Of Radiology*. 2015; 84(2): 235-41.
24. Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): A review of the evidence for use as a screening tool. *Clinical Radiology*. 2016; 71(2): 141-50.
25. Rafferty EA, Durand MA, Conant EF, et al. Breast cancer screening using tomosynthesis and digital mammography in dense and nondense breasts. *JAMA*. 2016; 315(16): 1784-6.
26. Mercier J, Kwiatkowski F, Abrial C, Boussion V, Dieu-De Fraissinette V, Marraoui W, et al. The role of tomosynthesis in breast cancer staging in 75 patients. *Diagnostic and Interventional Imaging*. 2015; 96(1): 27-35.
27. Mun HS, Kim HH, Shin HJ, Cha JH, Ruppel PL, Oh HY, et al. Assessment of extent of breast cancer: Comparison between digital breast tomosynthesis and full-field digital mammography. *Clinical radiology*. 2013; 68(12): 1254-9.
28. Seo N, Kim HH, Shin HJ, Cha JH, Kim H, Moon JH, et al. Digital breast tomosynthesis versus full-field digital mammography: comparison of the accuracy of lesion measurement and characterization using specimens. *Acta Radiol*. 2014; 55(6): 661-7.
29. Haq R, Lim YY, Maxwell AJ, Hurley E, Beetles U, Bundred S, et al. Digital breast tomosynthesis at screening assessment: Are two views always necessary? *British Journal of Radiology*. 2015; 88(1055).
30. Lång K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study. *European Radiology*. 2016; 26(1): 184-90.

Figure and Table legends:

Figure 1: Receiver Operating Characteristic (ROC) curves, showing accuracy of screening mammogram plus single-view DBT (experimental method) and screening mammogram plus supplementary mammographic views (standard method)

Table 1: Measurement agreement with histology in unifocal malignant lesions